An Introduction to Saddlepoint Methods

Elvezio Ronchetti

Research Center for Statistics
and
Geneva School of Economics and Management
University of Geneva, Switzerland

Elvezio.Ronchetti@unige.ch

http://www.unige.ch/ses/metri/ronchetti/

♦ Outline

- Introduction
- Saddlepoint Techniques
- Saddlepoint Approximation of the Density of the Maximum Likelihood Estimator in an Exponential Family
- Saddlepoint Approximation of the Density of M-estimators
- Tail Probabilities
- Marginal Distributions
- Small Sample Asymptotics and Robust Statistics

- **Saddlepoint Test**

- **Relationship with Nonparametric Techniques**

- **Saddlepoint Test: Applications**

- **Conclusions**

♦ Introduction

General Problem

Tail probabilities $P[T_n > t]$ are needed to carry out statistical inference (tests and confidence intervals).

Unless T_n has a simple form (e.g. linear)

and/or

the underlying distribution of the observations has a particular form (e.g. normal), tail probabilities cannot be computed exactly.

$\longrightarrow$ rely on asymptotic approximations

Unfortunately the asymptotic (normal) distribution can be a poor approximation of tail areas especially for moderate to small sample sizes or far out in the tails.

This is exactly the region of interest for constructing confidence intervals and tests.

Edgeworth Expansions

Can try to improve the accuracy by using e.g. Edgeworth expansions; cf. Feller(1971), Ch. 16.
They are obtained by a Taylor expansion of the characteristic function of the statistic of interest around 0, i.e. at the center of the distribution, followed by a Fourier inversion.

This leads to expansions of the distribution in powers of $n^{-1/2}$, where the leading term is the normal density.

By construction Edgeworth expansions provide in general a good approximation in the center of the density, but they can be inaccurate in the tails, where they can even become negative.

Saddlepoint Techniques

Saddlepoint techniques overcome these problems. They can be traced back to
Riemann(1892): method of steepest descent

Introduced into statistics by
Daniels(1954), *Ann. Math. Stat.*

These approximations exhibit a
relative error of order $O(n^{-1})$
to be compared with *absolute* errors of order $O(n^{-1/2})$ obtained by using Edgeworth expansions and similar techniques.

They provide very accurate numerical approximations for densities and tail areas down to small sample sizes and /or out in the tails.

General references (books)

- Field, C., Ronchetti, E.(1990)
  Small Sample Asymptotics,
  Institute of Mathematical Statistics Monograph Series, Hayward (CA).

- Jensen, J.L.(1995)
  Saddlepoint Approximations,
  Oxford University Press.

- Kolassa, J.E.(1997)
  Series Approximation Methods in Statistics,
  Springer, New York.

- Butler, R.W. (2007) *Saddlepoint Approximations with Applications*,
  Cambridge University Press.

General references (cont.)

- Brazzale, A., Davison, A. C., Reid, N. (2007)
  Applied Asymptotics: Case Studies in Small-Sample Statistics,
  Cambridge University Press.

- Ronchetti, E. (1997)
  "Introduction to Daniels(1954)",
  in *Breakthroughs in Statistics*, Vol. III,
  eds. S. Kotz and N.L. Johnson,
  Springer, New York, 171-176.

- Huber, P.J., Ronchetti, E.(2009)
  Robust Statistics,
  2nd edition, Wiley, New York,
  Chapter 14.

♦ Saddlepoint Techniques

For simplicity we derive the saddlepoint approximation of the density of the mean of n iid random variables.

However, it is more useful to derive accurate approximations in finite samples for *robust* statistics rather than for non-robust statistics like the mean
because errors due to deviations from the underlying model dominate errors due to finite sample approximations;
cf. Ronchetti & Ventura (2001)
Statistics & Computing

Therefore, later we focus on saddlepoint approximations for $M-$estimators.

$X_1, \cdots, X_n$ iid random variables from a distribution F on sample space $\mathcal{X}$.

$M(\lambda) = E[e^{\lambda X}]$: moment gener. fct.

$K(\lambda) = log\ M(\lambda)$: cumulant gener. fct.

Density $f_n(t)$ of the arithmetic mean (Fourier inversion):

$$
\begin{aligned}
f_n(t) &= \frac{n}{2\pi} \int_{-\infty}^{\infty} M^n(i\tau) e^{-in\tau t} d\tau \\
&= \frac{n}{2\pi i} \int_{\mathcal{I}} M^n(z) e^{-nzt} dz \\
&= \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} M^n(z) e^{-nzt} dz \\
&= \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} exp\{n[K(z) - zt]\} dz \ ,
\end{aligned}
$$

where $\mathcal{I}$ is the imaginary axis and $\tau \in \mathbb{R}$.

Choose $\tau = z_0$,

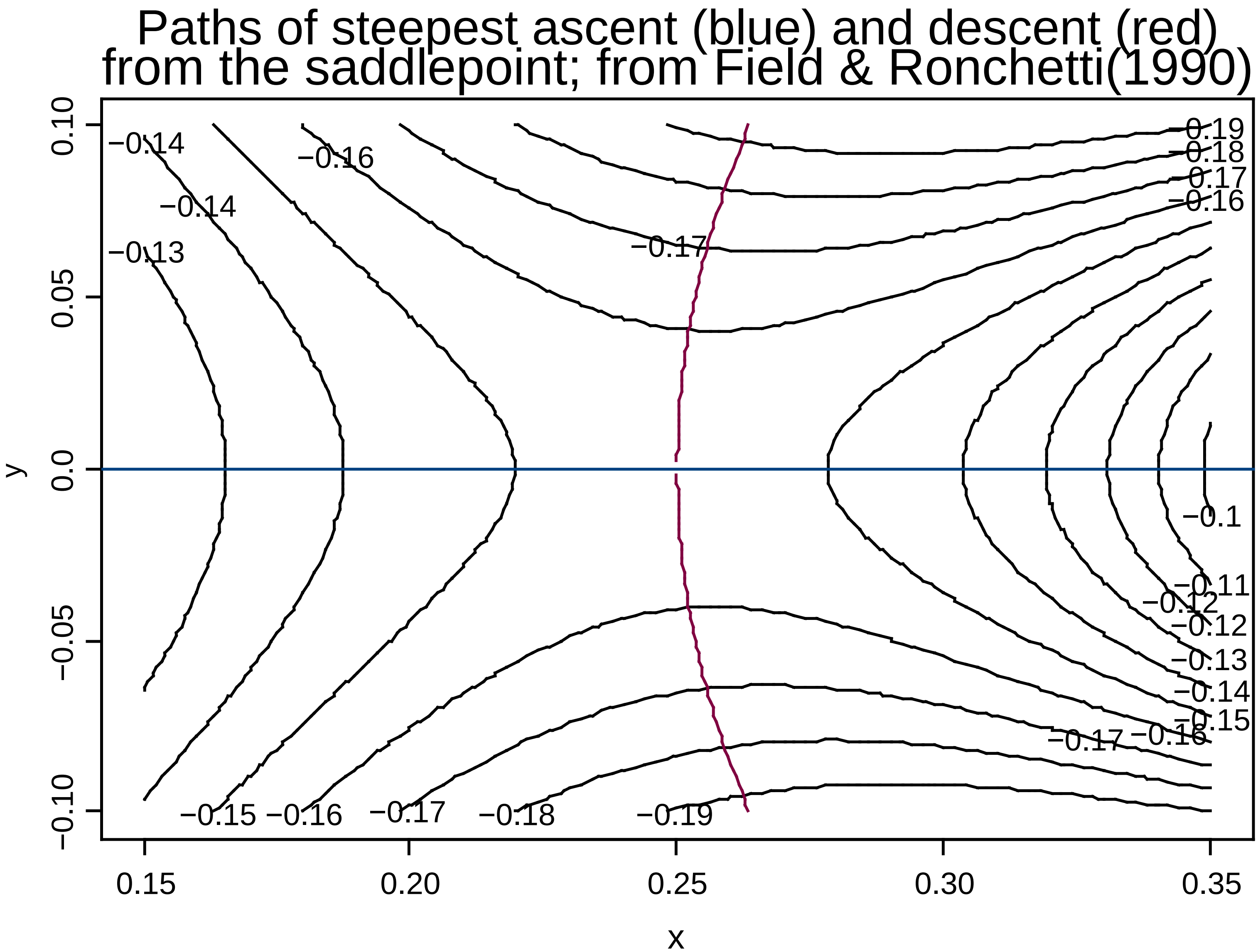
the saddlepoint of $w(z;t) = K(z) - zt$, i.e

$$\frac{\partial}{\partial z} w(z_0;t) = K'(z_0) - t = 0.$$

(Can prove $z_0 \in \mathbb{R}$)

Modify the integration path to go through the path of steepest descent (defined by $\mathcal{I}w(z;t) = 0$) from the saddlepoint z_0.

This captures most of the mass around the saddlepoint. The contributions to the integral outside a neighborhood of the saddlepoint are negligible.



Level curves and path of steepest descent from the saddlepoint $z_0 = .25$ for the surface $u(x, y) = \mathcal{R}w(z)$,
$w(z) = -\beta log(1 - \frac{z}{\alpha}) - zt$, $t = 2, \alpha = \beta = .5$
(mean of n iid variables from a Gamma distribution)


Monteviasco(I)1ora
Vezio 1ora 30 min.
Arosio 2ore 5 min.

This leads to the saddlepoint approximation $g_n(t)$, Daniels(1954), *Ann. Math. Stat.*

$$f_n(t) = g_n(t)\{1 + O(n^{-1})\},$$

where

$$g_n(t) = \left(\frac{n}{2\pi K''(\lambda(t))}\right)^{1/2} exp\{n[K(\lambda(t)) - \lambda(t)t]\}$$

and $\lambda(t)$ (saddlepoint) is the solution of

$$K'(\lambda) - t = 0.$$

The saddlepoint approximation $g_n(t)$ of $f_n(t)$ has relative error of order $O(n^{-1})$:

$$rel.\ err. = \frac{g_n(t) - f_n(t)}{f_n(t)} = O(n^{-1})$$

Alternative way to obtain the saddlepoint approximation is to use the idea of conjugate density, cf. Esscher (1932).

(1) First recenter the underlying density f at the point t where we want to evaluate the density of the mean,
i.e. define the conjugate density

$$f_t(x) = C(t) exp\{\alpha(t)(x-t)\} f(x),$$

where $C(t)$ and $\alpha(t)$ are chosen such that $f_t(x)$ is a density (it integrates to 1) and has expectation t.

f_t is the closest distribution to f in the (backward) Kullback-Leibler distance with expectation t, i.e. it minimizes

$$d_{KL}(g, f) = \int g(x) \log \left[\frac{g(x)}{f(x)}\right] dx,$$

under the conditions

$$g(x) \geq 0, \quad \int g(x) dx = 1, \quad \int x g(x) dx = t.$$

(2) Now use locally a normal approximation to the density of the mean based on the conjugate density f_t rather than f.

This is very accurate because under the conjugate density, we are approximating a density at the center at its expected value.

(3) The final step is to relate the density of the mean computed under the conjugate, say $f_{n,t}$, to the desired density f_n:

$$f_n(t) = C^{-n}(t) f_{n,t}(t).$$

This procedure is repeated for each point t and the conjugate density changes as we vary t.

It turns out that centering at t the conjugate density is equivalent to solving the saddle-point equation and the two approaches yield the same approximation, where
$-log\ C(t) = K(\lambda(t)) - \lambda(t)t$, $\lambda(t) = \alpha(t)$, and $K''(\lambda(t)) = \sigma^2(t)$, the variance of the conjugate density.

(Very) special case:

$F = \mathcal{N}(0,1); \quad \bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})$

$$\begin{aligned}
K(\lambda) &= \frac{\lambda^2}{2} \quad K''(\lambda(t)) = 1 \\
\lambda(t) &= t \\
-\log C(t) &= K(\lambda(t)) - \lambda(t) \cdot t = -\frac{1}{2}t^2 \\
C(t) &= \exp\{\frac{1}{2}t^2\} \\
g_n(t) &= \left(\frac{n}{2\pi}\right)^{1/2} \exp\{-\frac{1}{2}nt^2\}
\end{aligned}$$

Hampel(1973) who coined the expression *small sample asymptotics* to indicate the spirit of these techniques, proposed to re-center the original distribution combined with the expansion of the logarithmic derivative

$$f_n'/f_n$$

rather than the density f_n itself.

$\longrightarrow$ normalizing constant, i.e. the constant that makes the total mass equal to 1, must be determined numerically.

Advantage, since this rescaling improves further the approximation
order of the relative error of the approximation goes from $O(n^{-1})$ to $O(n^{-3/2})$.
Finally, this amounts to drop the constant $(n/2\pi)^{1/2}$ provided by the asymptotic normal distribution in the saddlepoint approximation and to renormalize the approximation, i.e.

$$\begin{aligned} g_n(t) &= c_n exp\{n[K(\lambda(t)) - \lambda(t)t]\}(K''(\lambda(t)))^{-1/2} \\ &= c_n C^{-n}(t)\sigma(t)^{-1}, \end{aligned}$$

where c_n is the normalizing constant, i.e the constant that makes the total mass $\int g_n(t)dt$ equal to 1.

Renormalization

Take values in the range
$t - \mu = O(n^{-1/2}),\ \mu = E[X_i]$

$$f_n(t) = g_n(t)\{1 + \frac{a(t)}{n} + O(n^{-2})\},$$

$$\begin{aligned} g_n(t) &= f_n(t)\{1 - \frac{a(t)}{n} + O(n^{-2})\} \\ &= f_n(t)\{1 - \frac{a(\mu)}{n} - \frac{(t-\mu)a'(\mu)}{n} \\ &+ O((t-\mu)^2/n) + O(n^{-2})\} \end{aligned}$$

$$
\begin{aligned}
c_n &= \int g_n(t)\,dt \\
&= 1 - \frac{a(\mu)}{n} + O(n^{-2})
\end{aligned}
$$

$$
\begin{aligned}
\frac{g_n(t)}{c_n} &= f_n(t)\frac{1 - \frac{a(\mu)}{n} + O(n^{-3/2})}{1 - \frac{a(\mu)}{n} + O(n^{-2})} \\
&= f_n(t)\{1 + O(n^{-3/2})\}
\end{aligned}
$$

For the mean:

$$
\begin{aligned}
a(\mu) &= -\frac{5}{24}\kappa_3^2 + \frac{1}{8}\kappa_4, \\
\kappa_3 &= E\left[\frac{X-\mu}{\sigma}\right]^3 \quad (skewness) \\
\kappa_4 &= E\left[\frac{X-\mu}{\sigma}\right]^4 - 3 \quad (kurtosis)
\end{aligned}
$$

♦ Saddlepoint Approximation of the Density of the Maximum Likelihood Estimator in an Exponential Family

$$X_1, \ldots, X_n \quad \text{iid}$$

$$X_i \sim f_\theta(x) = e^{\theta^T A(x) - B(\theta) + D(x)}$$

MLE $\widehat{\theta}$: $B'(\widehat{\theta}) = \frac{1}{n} \sum_{i=1}^n A(X_i) (= \bar{A})$

(1) Find the SAD approximation of the density of the mean $\bar{A}$.

(2) Obtain the SAD approximation of the density of $\widehat{\theta}$ by transformation.

(1)

$$
\begin{aligned}
M_A(\lambda) &= \int e^{(\theta+\lambda)^T A(x) - B(\theta) + D(x)} dx \\
&= e^{B(\theta+\lambda) - B(\theta)} \\
K_A(\lambda) &= B(\theta+\lambda) - B(\theta)
\end{aligned}
$$

Saddlepoint λ_0: $K'_A(\lambda_0) = \bar{A}$
$\rightarrow \lambda_0 = \hat{\theta} - \theta$

SAD approx. to the density of $\bar{A}$:

$$
\begin{aligned}
& \tilde{f}_n(\bar{A}; \theta) \\
= & \left[\frac{n}{2\pi |B''(\hat{\theta})|} \right]^{1/2} e^{n(B(\hat{\theta}) - B(\theta) - (\hat{\theta}-\theta)^T B'(\hat{\theta}))} \\
\cdot & \{1 + O(n^{-1})\}
\end{aligned}
$$

(2)

$B'(\hat{\theta}) = \bar{A}$
Jacobian of the transformation:
$|B''(\hat{\theta})| d\hat{\theta} = d\bar{A}$

SAD approx. to the density of the MLE $\widehat{\theta}$:

$$
\begin{aligned}
& f_n(\widehat{\theta};\theta) \\
= \; & \left[\frac{n|B''(\widehat{\theta})|}{2\pi}\right]^{1/2} e^{n(B(\widehat{\theta})-B(\theta)-(\widehat{\theta}-\theta)^T B'(\widehat{\theta}))} \\
\cdot \; & \{1+O(n^{-1})\} \\
= \; & \left[\frac{n}{2\pi}\right]^{1/2} \frac{e^{nB(\widehat{\theta})-n\widehat{\theta}^T B'(\widehat{\theta})}}{e^{nB(\theta)-n\theta^T B'(\widehat{\theta})}} |B''(\widehat{\theta})|^{1/2} \\
\cdot \; & \{1+O(n^{-1})\}
\end{aligned}
$$

$$f_n(\widehat{\theta};\theta) = c_n \frac{L(\theta)}{L(\widehat{\theta})} |j(\widehat{\theta})|^{1/2} \{1 + O(n^{-1})\}$$

where $L(\cdot)$ is the likelihood function and $j(\theta)$ the Fisher information.

Daniels in the discussion of
Cox(1958), *J. Roy. Stat. Soc. B* ;
Barndorff-Nielsen(1983), *Biometrika*

⧫ Saddlepoint Approximation of the Density of M-estimators

$X_1, \cdots, X_n$ iid random vectors from a distribution F on sample space $\mathcal{X}$.

M-functional $\beta(F) \in R^q$:

$$E_F\{\psi(X;\beta)\} = 0$$

M-estimator T_n of β:

$$\sum_{i=1}^{n} \psi(X_i; T_n) = 0.$$

$$f_{T_n}(t) = c_n e^{n \overbrace{K_\psi(\lambda(t);t)}^{=-h(t)}} . \quad \mid B(t) \| \Sigma(t) \mid^{-1/2} \left(1 + O(n^{-1})\right)$$

$\lambda(t)$ (saddlepoint) :

$$\begin{aligned} \frac{\partial}{\partial \lambda} K_\psi(\lambda;t) &= 0 \\ i.e. \quad E\{\psi(X;t) e^{\lambda^T \psi(X;t)}\} &= 0 \\ i.e. \quad E_t\{\psi(X;t)\} &= 0 \end{aligned}$$

$$\begin{aligned} K_\psi(\lambda;t) &= \log E\{e^{\lambda^T \psi(X;t)}\} \\ B(t) &= E_t\{-\frac{\partial \psi(X;t)}{\partial t}\}, \\ \Sigma(t) &= E_t\{\psi(X;t)\psi^T(X;t)\}, \end{aligned}$$

Field & Hampel (1982) *Biometrika*

Field (1982) *Ann. Stat.*

Almudevar, Field, Robinson (2000) *Ann. Stat.*

- E_t is the expectation taken with respect to the conjugate density $f_t(x) = C(t) exp\{\lambda^T(t)\psi(x;t)\}f(x)$ (conjugate density for the linearized version of the $M-$estimator).

- The error term holds uniformly for all t in a compact set.

- Still most general (and accurate) second-order formula!

- Existence of cum. gen. fct. $\longleftrightarrow$ Robustness

Numerical example:
Saddlepoint approximation of the
Huber estimator
when the underlying distribution is Cauchy

The following table gives

$\%\ relative\ err. = 100(saddl.appr. - exact)/exact$

of the saddlepoint approximation of upper tail areas $P[T_n > t]$ for the Huber estimator, i.e. an $M-$estimator with

$$\begin{aligned}\psi(x;t) &= x-t \quad if \mid x-t \mid \leq c \\ &= c \cdot sgn(x-t)\ otherwise.\end{aligned}$$

Exact tail area: numerical integration of the density obtained by fast Fourier transform (A. Marazzi)
Saddlepoint approximation obtained by numerical integration of the saddlepoint density approximation. Direct saddlepoint approximations of tail areas are also available.

t n	1	2	3	4	5	6	7
1	-12.3	8.0	-4.4	0.8	-1.5	0.6	-0.7
3	-21.0	23.3	-12.6	14.1	-7.0	8.5	-4.0
5	-33.6	33.6	-24.9	24.9	-16.2	18.6	-12.2
7	-43.5	40.3	-37.2	33.1	-28.0	27.8	-16.7
9	-51.2	44.8	-47.8	38.6	-37.5	35.7	-29.8

% relative err. of the saddl. approx. for tail areas of the Huber estimator ($c = 1.4$) for the Cauchy underlying distribution;
from Field and Hampel(1982), *Biometrika*.

The errors are under control even in the extreme tails.
For instance for $n = 7$ and $t = 9$ (relative err. 30%), the actual difference is .99995-.99994 and the approximation is usable at the .005% level.

♦ Tail Probabilities

Convenient to have direct approx. of tail probabilities

For the mean:

$$
\begin{aligned}
\bar{F}_n(t) &= P[\bar{X}_n > t] \\
&= \int_t^\infty \frac{n}{2\pi}\int_{-\infty}^{\infty} M^n(i\tau)e^{-in\tau s}d\tau ds \\
&= \int_t^\infty \frac{n}{2\pi}\int_{-\infty}^{\infty} exp\{n[K(i\tau)-i\tau s]\}d\tau ds\ .
\end{aligned}
$$

Reverse the order of integration and evaluate the integral with respect to s:

$$
\begin{aligned}
\bar{F}_n(t) &= P[\bar{X}_n > t] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} exp\{n[K(i\tau)-i\tau t]\}d\tau/i\tau \\
&= \frac{1}{2\pi i}\int_{\mathcal{I}} exp\{n[K(z)-zt]\}dz/z.
\end{aligned}
$$

Use method of steepest descent by taking into account that the function to be integrated has a pole in $z = 0$.
Make a change of variable from z to w such that

$$K(z) - zt = \frac{1}{2}w^2 - \gamma w,$$

where $\gamma = sgn(\lambda)(2(\lambda t - K(\lambda)))^{1/2}$,
$w = \gamma$ is the image of the saddlepoint $z = \lambda(t)$, and the origin is preserved.

$$P[\bar{X}_n > t] = \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} exp\{n[\frac{1}{2}w^2 - \gamma w]\}G_0(w)\frac{dw}{w},$$

where $G_0(w) = \frac{w}{z}\frac{dz}{dw}$.
This operation takes the term to be approximated from the exponent, where the errors can become very large, to the main part of the integrand.

$G_0(w)$ has removable singularities at $w = 0$ and $w = \gamma$ and can be approximated by a linear function $a_0 + a_1 w$, where
$a_0 = \lim_{w\to 0} G_0(w) = 1$ and
$a_1 = (\lim_{w\to\gamma} G_0(w) - 1)/\gamma = -\frac{1}{\gamma} + \frac{1}{\lambda(K''(\lambda))^{1/2}}$.

The integrals can now be evaluated analytically and by using again the notation

$$\begin{aligned}\gamma &= sgn(\lambda(t))(2(\lambda(t)t - K(\lambda(t))))^{1/2} \\ &= \sqrt{2\ \log C(t)},\end{aligned}$$

this leads to the following tail area approximation:

$$\begin{aligned}\bar{F}_n(t) &= P_F[T_n > t] \\ &= 1 - \Phi\left(\sqrt{2n\ \log C(t)}\right) \\ &+ \frac{C(t)^{-n}}{\sqrt{2\pi n}}\left(\frac{1}{\sigma(t)\lambda(t)} - \frac{1}{\sqrt{2\ \log C(t)}}\right) \\ &\cdot \left(1 + O(n^{-1})\right).\end{aligned}$$

where $\lambda(t)$, $\sigma^2(t) = \Sigma(t)$ are defined above and $C(t) = e^{-K_\psi(\lambda(t);t)} = (E\{e^{\lambda(t)^T\psi(X;t)}\})^{-1}$

Lugannani and Rice (1980), *Adv. Appl. Prob.* for the mean, $\psi(x;t) = x - t$,

Daniels(1983), *Biometrika* for location M-estimators.

n	t	Exact	Integr. SP	Tail Area SP
1	0.1	.46331	.46229	.46282
	1.0	.17601	.18428	.18557
	2.0	.04674	.07345	.07082
	2.5	.03095	.06000	.05682
	3.0	.02630	.05520	.05190
5	0.1	.42026	.42009	.42024
	1.0	.02799	.02799	.02799
	2.0	.00414	.00413	.00416
	2.5	.00030	.00043	.00043
	3.0	.00018	.00031	.00031
9	0.1	.39403	.39393	.39399
	1.0	.00538	.00535	.00537
	2.0	.000018	.000018	.000018
	2.5	.000004	.000005	.000005
	3.0	.000002	.000003	.000003

Tail probabilities of Huber's M-estimator with $c = 1.5$ when the underlying distribution is a 5% contaminated normal.
"Integr. SP" is obtained by numerical integration of the saddlepoint approximation to the density;
from Daniels(1983), *Biometrika*

n	t	Exact	Integr. SP	Tail Area SP
1	1	.25000	.28082	.28197
	3	.10242	.12397	.13033
	5	.06283	.08392	.09086
	7	.04517	.06484	.07210
	9	.03522	.05327	.06077
5	1	.11285	.11458	.11400
	3	.00825	.00883	.00881
	5	.00210	.00244	.00244
	7	.00082	.00105	.00104
	9	.00040	.00055	.00055
9	1	.05422	.05447	.05427
	3	.00076	.00078	.00078
	5	.000082	.000088	.000088
	7	.000018	.000021	.000021
	9	.000006	.000006	.000007

Tail probabilities of Huber's M-estimator with $c = 1.5$ when the underlying distribution is Cauchy.
"Integr. SP" is obtained by numerical integration of the saddlepoint approximation to the density;
from Daniels(1983), *Biometrika*

♦ Marginal Distributions

Often interested in marginal densities and tail probabilities of a single component, say the last one.
This requires the integration of the joint density with respect to the other components.
Apply Laplace's method to

$$\int f_{T_n}(t)dt_1 \cdots dt_{q-1} =$$
$$\int c_n exp\{nK_\psi(\lambda(t);t)\} \mid B(t) \| \Sigma(t) \mid^{-1/2} dt_1 \cdot \cdot dt_{q-1}$$
$$\left(1 + O(n^{-1})\right)$$

DiCiccio, Field, Fraser(1990), *Biometrika*
Tingley & Field(1990), *JASA*
Daniels & Young(1991), *Biometrika*
Wang(1993), *J. Appl. Prob.*
Jing & Robinson(1994), *Ann. Stat*
Fan & Field(1995), *Can. J. Stat.*
Davison, Hinkley, Worton(1995), *Stat.&Comp.*
Gatto & Ronchetti(1996), *JASA*

♦ Some Selected Applications

- Engineering
  signal detection
  Helstrom and Ritcey (1984), *IEEE Tr. Aer. Elec. Sys.*

- Biostatistics

  - Survival times in flowgraph models
    Butler and Bronson (2002), *JRSS B*
    Yau and Huzurbazar (2002), *Stat. in Med.*
    Lô, Heritier, Hudson (2009), *Comp. Stat. & Data An.*

  - Kolassa (2003), *Stat. Meth. in Med. Res.*

- Economics
  - insurance
    Embrechts *et al.* (1985), *Adv. Appl. Prob.*
    Gatto (2004), *Astin Bull.*
  - information and entropy econometrics
    Imbens, Spady, Johnson (1998), *Econometrica*
  - credit risk
    Gordy (2002), *Journal of Banking and Finance*
  - trans. densities of Markov processes
    Aït-Sahalia and Yu (2006), *J. Econometrics*

♦ Small Sample Asymptotics and Robust Statistics

Use small sample asymptotics

(1) as a technique to obtain accurate tail probabilities for robust estimators and test statistics;

(2) as an analytic tool to investigate the robustness properties of statistical procedures and to develop new robust procedures.

Ex. of (2)

- Tail Area Influence Function
  IF on SSA approx. of $P_F[T_n > t]$
  Field & Ronchetti (1985)
  Ronchetti & Ventura (2001)
  García-Pérez (2003, 2006, 2008)

- Saddlepoint Test; see below

- Robust Divergence Estimators and Tests
  Toma & Leoni-Aubin(2010)
  Toma & Broniatowski(2011)

Other Approaches

- Exact Representation
  Jureckova (1999)

- Higher-order Improvements
  Bellio, Greco, Ventura (2008)
  La Vecchia, Ronchetti, Trojani (2012)

- Robust Bootstrap and Subsampling
  Salibian-Barrera & Zamar (2002)
  Camponovo, Scaillet, Trojani (2012)

♦ Saddlepoint Test

Test a hypothesis

$$u(\beta) = \eta_0 \in R^{q_1}, \qquad q_1 \leq q$$

Test statistic based on M-estimator T_n of β.

$q_1 = 1$
Can use saddlepoint approximations for marginal distribution of $u(T_n)$.

$q_1 > 1$
Use **one-dimensional test statistic** $h(u(T_n))$ whose distribution can be approximated with relative error $O(n^{-1})$.

Robinson, Ronchetti, Young (2003),
Ann. Stat

Simple Hypothesis

$H_0 : \beta = \beta_0 \in R^q$

$$\begin{aligned} p - value &= P_{H_0}[h_n(T_n) \geq h_n(t_n)] \\ &= (1 - Q_q(h_n(t_n)))(1 + O(n^{-1})) \end{aligned}$$

- $Q_q(.)$: cumul. distr. fct. of χ^2_q
- T_n $M-$est. and t_n its obs. value
- $h_n(y) = 2nh(y) = 2n \ \sup_\lambda\{-K_\psi(\lambda; y)\}$
- $K_\psi(\lambda; \beta) = \log E\{e^{\lambda^T \psi(X;\beta)}\}$
  cumulant generating fct. of $\psi(X; \beta)$

Sketch of the proof

(1) Density of M-estimator

$$f_{T_n}(t) = c_n e^{n \overbrace{K_\psi(\lambda(t);t)}^{=-h(t)}} \cdot \quad | B(t) \| \boldsymbol{\Sigma}(t) |^{-1/2} \left(1 + O(n^{-1})\right)$$

$\lambda(t)$ (saddlepoint) : $\frac{\partial}{\partial \lambda} K_\psi(\lambda;t) = 0$

$$B(t) = O(1) \qquad \boldsymbol{\Sigma}(t) = O(1)$$

Field & Hampel (1982) *Biometrika*
Field (1982) *Ann. Stat.*
Almudevar, Field, Robinson (2000) *Ann. Stat.*

(2) Integration and transformation

p-value

$$= \int_A c_n e^{-nh(y)} \mid B(y) \mid\mid \Sigma(y) \mid^{-1/2} (1 + 0(n^{-1}))dy$$

$$= \int_{\tilde{A}} c_n e^{-nh(zn^{-1/2})} \mid B(zn^{-1/2}) \mid\mid \Sigma(zn^{-1/2}) \mid^{-1/2}$$

$$\cdot n^{-1/2}(1 + O(n^{-1}))dz$$

$$\tilde{A} = \left\{ z \mid h(zn^{-1/2}) \geq h(t_n) \right\}$$

Two transformations :

$$z \underset{v(\text{polar transf.})}{\longrightarrow} p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \underset{\tilde{v}}{\longrightarrow} s = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

$$\begin{aligned} p_1 &= (z^T z)^{1/2} \qquad & s_1 \equiv w &= 2nh(n^{-1/2} v^{-1}(p)) \\ p_2 &= \ldots & s_2 &= p_2 \end{aligned}$$

Jacobians :

$$J_v = (z^T z)^{\frac{q-1}{2}} \qquad J_{\tilde{v}} = \frac{n^{-1/2}(z^T z)^{1/2}}{2h'(zn^{-1/2})^T z}$$

p-value

$$= \int_{h_n(t_n)}^{\infty} c_n e^{-\frac{w}{2}} \left\{ \int_{S_q} \delta(w, s_2)(1 + O(n^{-1})) ds_2 \right\} dw$$

where

$$\begin{aligned} \delta(w, s_2) &\equiv \Delta(z) \\ &= n^{-1} \mid B(zn^{-1/2}) \mid\mid \Sigma(zn^{-1/2}) \mid^{-1/2} \\ &\quad \cdot \frac{(z^T z)^{q/2}}{2h'(zn^{-1/2})^T z} \end{aligned}$$

and S_q : surface of the q-dim. sphere

(3) Expand $\Delta(z)$ about $z = 0$

$$\begin{aligned}\Delta(z) \; &= \; \frac{1}{2} n^{-1/2} w^{q/2} \mid B(0) \mid\mid \Sigma(0) \mid^{-1/2} \\ &\cdot (1 + n^{-1/2} b(z) + O(n^{-1}))\end{aligned}$$

$b(z)$ is an odd fonction : $b(z) = -b(-z)$
$\Longrightarrow \int\limits_{S_q} b(z) ds_2 = 0$

$\Longrightarrow$ **term** $0(n^{-1/2})$ **disappears !**

Discussion

(1) $h(y) = sup_\lambda\{-K_\psi(\lambda; y)\} = -K_\psi(\lambda(y); y)$

where $\lambda(y)$ is the saddlepoint satisfying

$$\frac{\partial}{\partial\lambda}K_\psi(\lambda; y) = 0.$$

– $h(y)$ is the **exponent** in the saddlepoint approx. of the density of $M-$est.

– essential quantity for **Laplace approx.** of integrals to marginalize

– when $\psi(x; \beta) = x - \beta$,
$h(\cdot)$ is the Legendre transform of the cumulant generating function of X

(2) The approx. has **relative error** $O(n^{-1})$ in the normal region $\hat{u} = O(n^{-1/2})$, where $\hat{u} = [2h(t_n)]^{1/2}$

To get this result in the **large deviations region** $\hat{u} < C$, we can get a "Lugannani-Rice adjustment"

$$
\begin{aligned}
p - value &= 1 - Q_q(n\hat{u}^2) \\
&= +\frac{1}{n} c_n \hat{u}^q \; e^{-n\frac{\hat{u}^2}{2}} \left[\frac{G(\hat{u}) - 1}{\hat{u}^2}\right] \\
&= +(1 - Q_q(n\hat{u}^2)) \; O(n^{-1})
\end{aligned}
$$

where $\hat{u} = [2h(t_n)]^{1/2}$, $n\hat{u}^2 = h_n(t_n)$, $c_n = n^{q/2}/(2^{q/2}\Gamma(q/2))$.

Since $\frac{G(\hat{u})-1}{\hat{u}^2}$ is bounded for $\hat{u} < C$, we get:

$$p - value = (1 - Q_q(n\hat{u}^2)) \; (1 + O(n^{-1} + \hat{u}^2))$$

(3) Expansion about $z = 0$ gives (w.l.o.g $h^{''}(0) = I$)

$$\begin{aligned} w &= 2n\ h(zn^{-1/2}) = h_n(zn^{-1/2}) \\ &= (z^T z)(1 + n^{-1/2} a(z) + O(n^{-1})) \end{aligned}$$

Test statistic $h_n(T_n)$ **asymptotically (first order) equivalent** to Wald and score test based on $M-$est. T_n.

In particular: **same (first order) robustness properties**;
cf. Heritier & Ronchetti(1994), *JASA*

- Relative error $O(n^{-1})$ for $h(.)$.
  Does not hold for Wald, score, likelihood ratio tests (classical or robust).
  $h(.)$ is a **"better scale" for approaching χ^2** and a better (pivotal) quantity to bootstrap.

(4) The saddlepoint test is the **log-likelihood ratio test** when T_n is the MLE and the model belongs to the **exponential family**; see below.

(5) (Very) special case:
$F = \mathcal{N}(0,1); \quad \bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})$

$K_\psi(\lambda; y) = \frac{\lambda^2}{2} - \lambda \cdot y$

$\frac{\partial}{\partial \lambda} K_\psi(\lambda; y) = \lambda - y = 0 \Rightarrow \lambda(y) = y.$

$$\begin{aligned} h(y) &= sup_\lambda \{-K_\psi(\lambda; y)\} = -K_\psi(\lambda(y); y) \\ &= sup_\lambda \{-\frac{\lambda^2}{2} + \lambda \cdot y\} = \frac{1}{2} y^2 \end{aligned}$$

$\Rightarrow 2nh(\bar{X}_n) = n\bar{X}_n^2$

(6) The result can be extended to the case where the $X_i's$ are not identically distributed (ex. regression).

(7) Test can be inverted to obtain confidence regions.

The Saddlepoint Test for Exponential Models is the Log-likelihood Ratio Test

$$X_1, \ldots, X_n \quad \text{iid}$$

$$X_i \sim f_\theta(x) = e^{\theta^T A(x) - B(\theta) + D(x)}$$

$H_0 : \theta = \theta_0$

$$\begin{aligned} \log f_\theta(x) &= \theta^T A(x) - B(\theta) + D(x) \\ \psi(x;\theta) &= \frac{\partial \log f_\theta(x)}{\partial \theta} = A(x) - B'(\theta) \end{aligned}$$

MLE $\widehat{\theta}$: $B'(\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^n A(X_i) (= \bar{A})$

$$
\begin{aligned}
K_\psi(\lambda;\theta) &= \log E_{F_{\theta_0}}\left[e^{\lambda^T\psi(X;\theta)}\right] \\
&= \log \int e^{\lambda^T\psi(x;\theta)} f_{\theta_0}(x)dx \\
&= \log \int e^{\lambda^T A(x)-\lambda^T B'(\theta)} \cdot e^{\theta_0^T A(x)-B(\theta_0)+D(x)} dx \\
&= -\lambda^T B'(\theta) - B(\theta_0) + \log \int e^{(\lambda+\theta_0)^T A(x)+D(x)} dx \\
&= -\lambda^T B'(\theta) - B(\theta_0) + B(\lambda+\theta_0)
\end{aligned}
$$

$$
\begin{aligned}
\lambda : K'_\psi(\lambda;\theta) = 0 \qquad &\Leftrightarrow \quad -B'(\theta) + B'(\lambda+\theta_0) = 0 \\
&\Leftrightarrow \quad \lambda = \theta - \theta_0
\end{aligned}
$$

$$
\begin{aligned}
h(\theta) &= \sup_{\lambda}\left[-K_{\psi}(\lambda;\theta)\right] = -K_{\psi}(\theta-\theta_0;\theta) \\
&= (\theta-\theta_0)^T B'(\theta) + B(\theta_0) - B(\theta)
\end{aligned}
$$

$$
\begin{aligned}
&\underline{2nh(\widehat{\theta})} \\
&= 2n\left\{\left(\widehat{\theta}-\theta_0\right)^T \underbrace{B'(\widehat{\theta})}_{=\bar{A}} + B(\theta_0) - B(\widehat{\theta})\right\} \\
&= 2\left(\widehat{\theta}^T \sum_{i=1}^{n} A(X_i) - nB(\widehat{\theta})\right) \\
&- 2\left(\theta_0^T \sum_{i=1}^{n} A(X_i) - nB(\theta_0)\right) \\
&= \underline{2\log\frac{L(\widehat{\theta})}{L(\theta_0)}}
\end{aligned}
$$

where $L(\cdot)$ is the likelihood function.

Composite Hypothesis

$H_0 : u(\beta) = \eta_0 \quad , \quad u : R^q \to R^{q_1}$

$$\begin{aligned} p - value &= P_{H_0}[h_n(u(T_n)) \geq h_n(u(t_n))] \\ &= (1 - Q_{q_1}(h_n(u(t_n))))(1 + O(n^{-1})) \end{aligned}$$

- $Q_{q_1}(.)$: cumul. distr. fct. of $\chi^2_{q_1}$
- T_n $M-$est. and t_n its obs. value
- $$\begin{aligned} h_n(y) &= 2nh(y) \\ &= 2n \inf_{\{\beta : u(\beta) = y\}} \sup_{\lambda} \{-K_\psi(\lambda; \beta)\} \end{aligned}$$
- $K_\psi(\lambda; \beta) = \log E\{e^{\lambda^T \psi(X;\beta)}\}$
  $h(.)$: Legendre transform of $K_\psi(.;.)$.

Sketch of the proof

1. q-dimensional joint density of T_n

$$\downarrow \quad \text{Laplace's method}$$

q_1-dimensional joint density of $u(T_n)$

$$f_{u(T_n)}(y) = \tau_n e^{-nh(y)} \gamma(y) \cdot (1 + O(n^{-1}))$$

2. Continue as in the case of a simple hypothesis

$$h(\cdot): \text{ profile } -K_\psi(\lambda(\beta); \beta)$$

♦ Relationship with Nonparametric Techniques

Saddlepoint approx. require the specification of the underlying distr. F of the observations.

F enters in the approximation only through the expected values defining $K_\psi(\lambda; t)$, $B(t)$, $\Sigma(t)$.
$\longrightarrow$ estimate F by its empirical distr. $\widehat{F}$
$\longrightarrow$ empirical (or nonparametric) saddlepoint approx.
In particular

$$\widehat{K}_\psi(\widehat{\lambda}; t) = \log\{n^{-1} \sum_{i=1}^{n} exp\{\widehat{\lambda}^T \psi(x_i; t)\},$$

where $\widehat{\lambda}(t)$, the empirical saddlepoint, is the solution of the equation

$$\sum_{i=1}^{n} \psi(x_i; t) exp\{\widehat{\lambda}^T \psi(x_i; t)\} = 0.$$

- Empirical saddlepoint approx.: alternative to bootstrap

- Resampling is replaced by the computation of the root of the empirical saddlepoint equation

- Error properties of these approx., see Ronchetti and Welsh (1994), *JRSS B*

  Sowell (2007, 2009)
  Holcblat (2015)

  Aeberhard, Cantoni, Heritier (2017)
  Comp. Stat. & Data An.

Connection Between Empirical Saddlepoint Approx. and Empirical Likelihood

Monti and Ronchetti (1993), *Biometrika*

$$n\hat{K}_{\psi}(\hat{\lambda};t) = -\frac{1}{2}\hat{W}(t) + \frac{1}{6}n^{-1/2}\Gamma(u) + O(n^{-1}),$$

where $u = n^{1/2}(t - T_n)$, T_n is the $M-$estimator defined by $\psi(\cdot;\cdot)$,

$$\hat{W}(t) = 2\sum_{i=1}^{n} log\{1 + \hat{\xi}(t)^T\psi(x_i;t)\}$$

is the empirical likelihood ratio statistic (Owen (1988), *Biometrika*), where $\hat{\xi}(t)$ satisfies

$$\sum_{i=1}^{n} \frac{\psi(x_i;t)}{1 + \hat{\xi}(t)^T\psi(x_i;t)} = 0.$$

Furthermore,

$$\Gamma(u) = -n^{-1}\sum_{i=1}^{n}\{u^T\hat{V}^{-1}\widehat{IF}(x_i;T,F)\}^3,$$

where

$$\widehat{IF}(x_i;T,F) = \hat{B}(T_n)^{-1}\psi(x_i;T_n)$$

is the empirical influence function of T_n,

$$\hat{V} = \hat{B}(T_n)^{-1}\Sigma(T_n)\{\hat{B}(T_n)^T\}^{-1}$$

is the estimated covariance matrix of T_n,

$$\begin{aligned}\hat{B}(T_n) &= n^{-1}\sum_{i=1}^{n}-\frac{\partial\psi(x_i;t)}{\partial t}\,|_{T_n},\\ \hat{\Sigma}(T_n) &= n^{-1}\sum_{i=1}^{n}\psi(x_i;T_n)\psi(x_i;T_n)^T.\end{aligned}$$

$\longrightarrow$ $-2n\hat{K}_\psi(\hat{\lambda};t)$ and $\hat{W}(t)$ are asymptotically (first order) equivalent

$\longrightarrow$ correction term for empirical likelihood ratio statistic to be equivalent to the empirical saddlepoint statistic up to order $O(n^{-1})$.

This correction term depends on the skewness of $\widehat{IF}(x;T,F)$ and in the univariate case,

$$\frac{1}{6}n^{-1/2}\Gamma(u) = -u^3\widehat{V}^{-3/2}a \ ,$$

where

$$a = \frac{1}{6}\sum_{i=1}^{n}\widehat{IF}(x_i;T,F)^3/\{\sum_{i=1}^{n}\widehat{IF}(x_i;T,F)^2\}^{3/2}$$

is the nonparametric estimator of the acceleration constant appearing in the BC_a method of Efron (1987), *JASA*, eqn (7.3), p.178.

Empirical Exponential Likelihood Tests

Distr. F unknown $\rightarrow$ nonparametric version
Basic idea:

- Define an empirical distr. $\widehat{F}_0$ with (exponential) weights $(p_1, \ldots, p_n)$ at the observations $(x_1, \ldots, x_n)$
  - closest (in Kullback-Leibler) distance to the empirical distr. with weights $p_i = \frac{1}{n}$
  - consistent with the null hypothesis $H_0 : u(\beta) = \eta_0$.

- Compute $K_\psi^0(\lambda; \beta)$, the empirical cumulant generating fct. of $\widehat{F}_0$.

- Compute the test statistic $h(.)$ and compare with quantiles of a χ^2 distr.

- Related to **exponential tilting**;
  cf.
  Efron(1982), *SIAM*
  DiCiccio&Romano(1992),*Int. Stat. Rev.*

  and **exponential empirical likelihood**
  cf.
  Monti & Ronchetti(1993), *Biometrika*
  Field, Robinson, Ronchetti(2007), *Ann. Inst. Stat. Math*

- Replace bootstrap resampling by optimization

♦ Saddlepoint Test: Applications

Example 1 RRY(2003)

$\psi(x;\theta) = x - \theta$, $q = 3$, $n = 20$

X_i is distributed as a vector of 3 independent exponential variables with means 1.

Elementary calculations give:

$$h(y) = \sum_{j=1}^{3} [(y_j - 1) - \log y_j]$$

In this case $T_n = \bar{X}_n$ and

$n\bar{X}_n$ is distributed as a vector of 3 independent Gamma variates with shape parameter n.

Generate 10,000 Monte Carlo replicates of $2nh(\bar{X}_n)$ and compare these to the approximating χ^2_3 distribution.

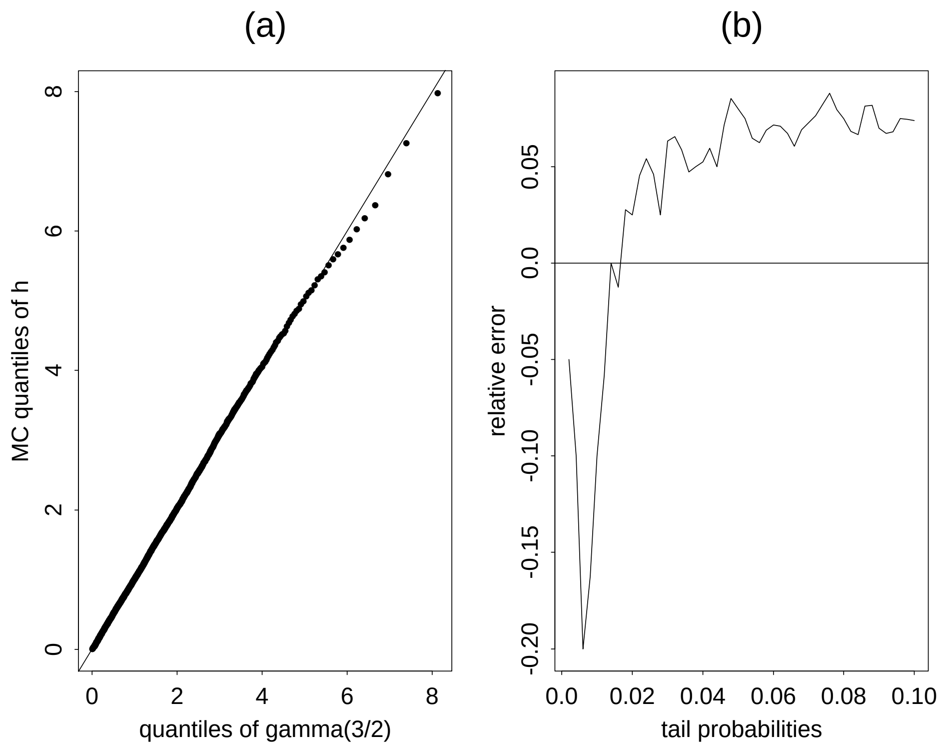


Figure 1 (a)

Q-Q plot of 10,000 Monte Carlo samples of $2nh(\bar{X}_n)$ with the theoretical quantiles (taking each 100th quantile in the plot)

Figure 1 (b)

Relative errors of the tail probabilities from 10,000 Monte Carlo trials compared to the χ^2_3 approximation.

The relative error is

$$(P(2nh(\bar{X}_n) > v_\alpha) - \alpha)/\alpha,$$

where $P(\chi^2_3 > v_\alpha) = \alpha$,

for $\alpha = .02, .04, \cdots, .1$.

Example 2 RRY(2003)

As in Example 1

$\psi(x;\theta) = x - \theta$, $q = 3$, $n = 20$

Draw a sample of size n from a 3 dimensional distribution of independent exponential variables with mean 1.

Obtain $\widehat{h}(\bar{x}_n)$ and for each of $10,000$ bootstrap samples from $\widehat{F}_0$ obtain $\widehat{h}(\bar{x}_n^*)$.

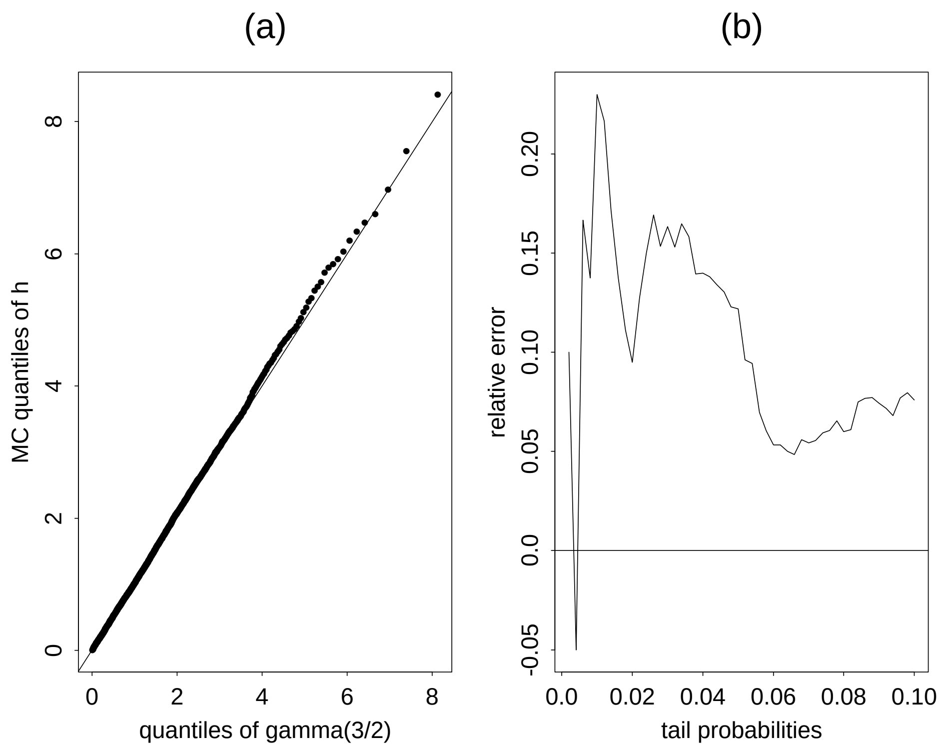


Figure 2(a)
Q-Q plot as in Example 1

Figure 2(b)
Relative error for tail areas of the
χ^2- approximation as

$$(P(2n\hat{h}(\bar{X}_n^*) \geq v_\alpha) - \alpha)/\alpha,$$

where $P(\chi_3^2 > v_\alpha) = \alpha$,
for $\alpha = .02, .04, \cdots, .1$

Example 3: Robust Regression RRY(2003)

$$y = x^T\theta + e,$$

where $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})$, $x = (1, x^{(2)}, x^{(3)})$
$x^{(2)}, x^{(3)}$ are independent $U[0,1]$
$H_0 : \theta^{(2)} = \theta^{(3)} = 0$
The errors e are from the distribution $(1-\epsilon)\Phi(t) + \epsilon\Phi(t/s)$ with different settings of ϵ, s. The M-estimator of T_n satisfies

$$\sum_{i=1}^{n} \psi(y_i; T_n) = 0,$$

where

$$\psi(y;\theta) = \psi_c\left(\frac{y - x^T\theta}{\sigma}\right)x,$$

for $\psi_c(r) = \min\{c, \max(-c, r)\}$ and $c = 1.5$. The scale parameter σ is estimated by Huber's Proposal 2.

Test statistics:

- empirical likelihood statistic $2n\widehat{h}(u(T_n))$
- robust Wald, score, and likelihood ratio test statistic

10,000 Monte Carlo samples of size $n = 20$
For the 25 values of $\alpha = 1/250, 2/250, \cdots,$ 25/250, we obtained the proportion of times out of 10,000 that the statistic, S_n say, exceeded v_α, where $P(\chi^2_2 \geq v_\alpha) = \alpha$.
For each Monte Carlo sample we obtained 299 bootstrap samples and calculated a bootstrap p-value, the proportion of the 299 bootstrap samples giving a value S_n^* of the statistic exceeding S_n.

The bootstrap test of nominal level α rejects H_0 if the bootstrap $p-$value is less than α.

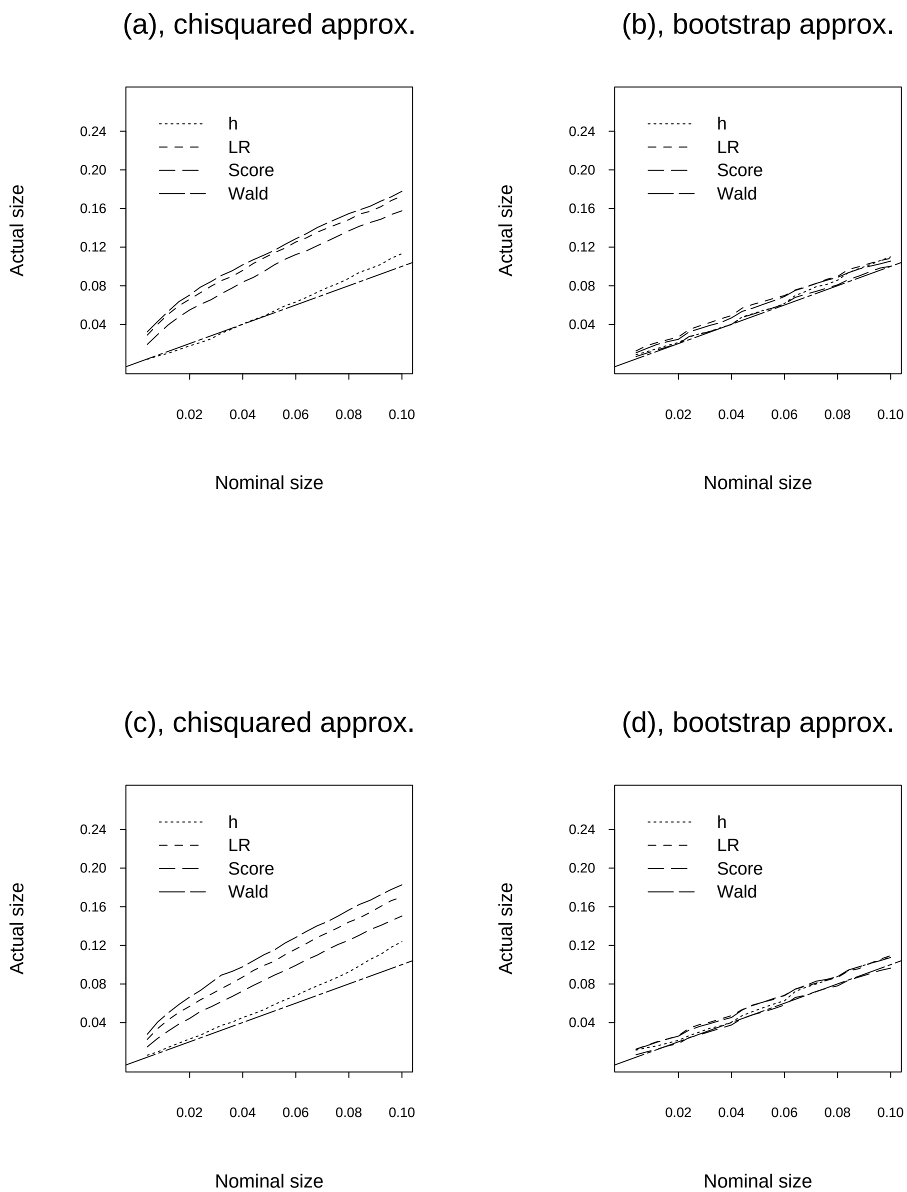


Figure 3
Actual size vs. nominal size for $\widehat{h}(u(T_n))$ and the three other tests based on χ^2 and bootstrap approx. (a), (b): $e \sim \Phi(.)$
(c), (d): $e \sim .99\Phi(.) + .01\Phi(./5)$

It is clear that the χ^2- approximation for $\widehat{h}(u(T_n))$ is much better than the corresponding χ^2- approximations for the other statistics.
Moreover, tests based on all the statistics are quite accurately approximated under the bootstrap.

Example 4: Robust GLM

Lô & Ronchetti (2009), *J. Mult. Anal.*

Example 5: Diffusion Models

Czellar & Ronchetti (2010), *Biometrika*

Geometric Brownian motion with drift:

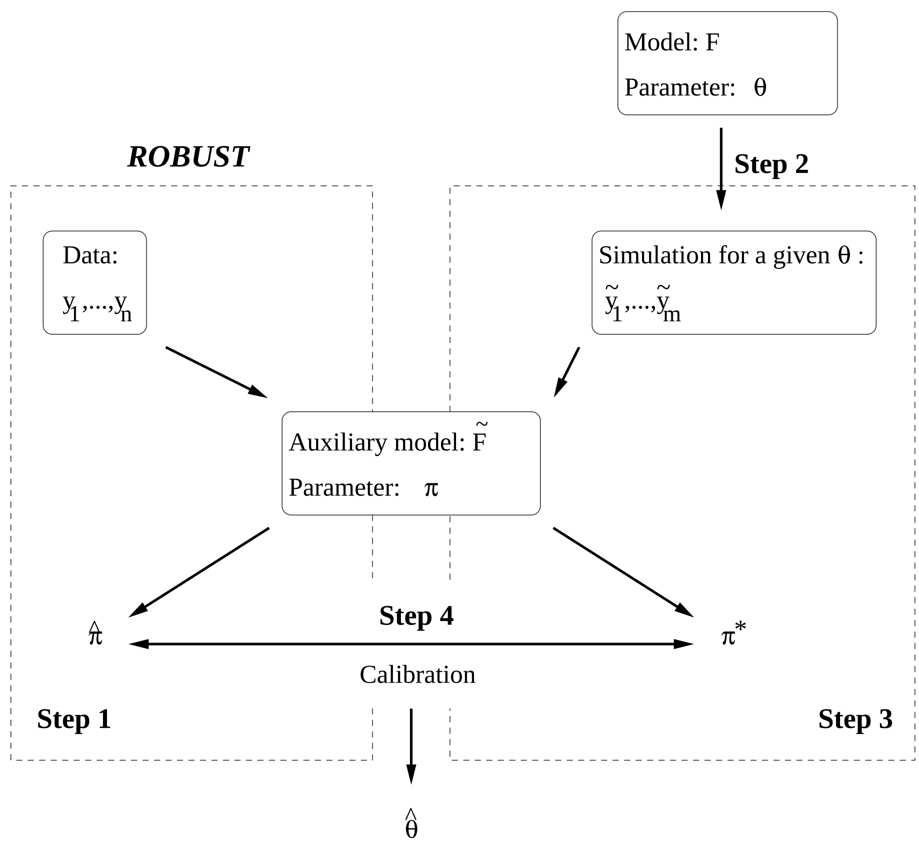
$$dy_t = \beta y_t + \sigma y_t dW_t \,,$$

where W_t is a standard Wiener process.

Estimation and testing by indirect inference using an auxiliary model obtained through a simple discretization:

$$y_t = (1 + \mu_1) y_{t-1} + \mu_2 y_{t-1} \epsilon_t \,,$$

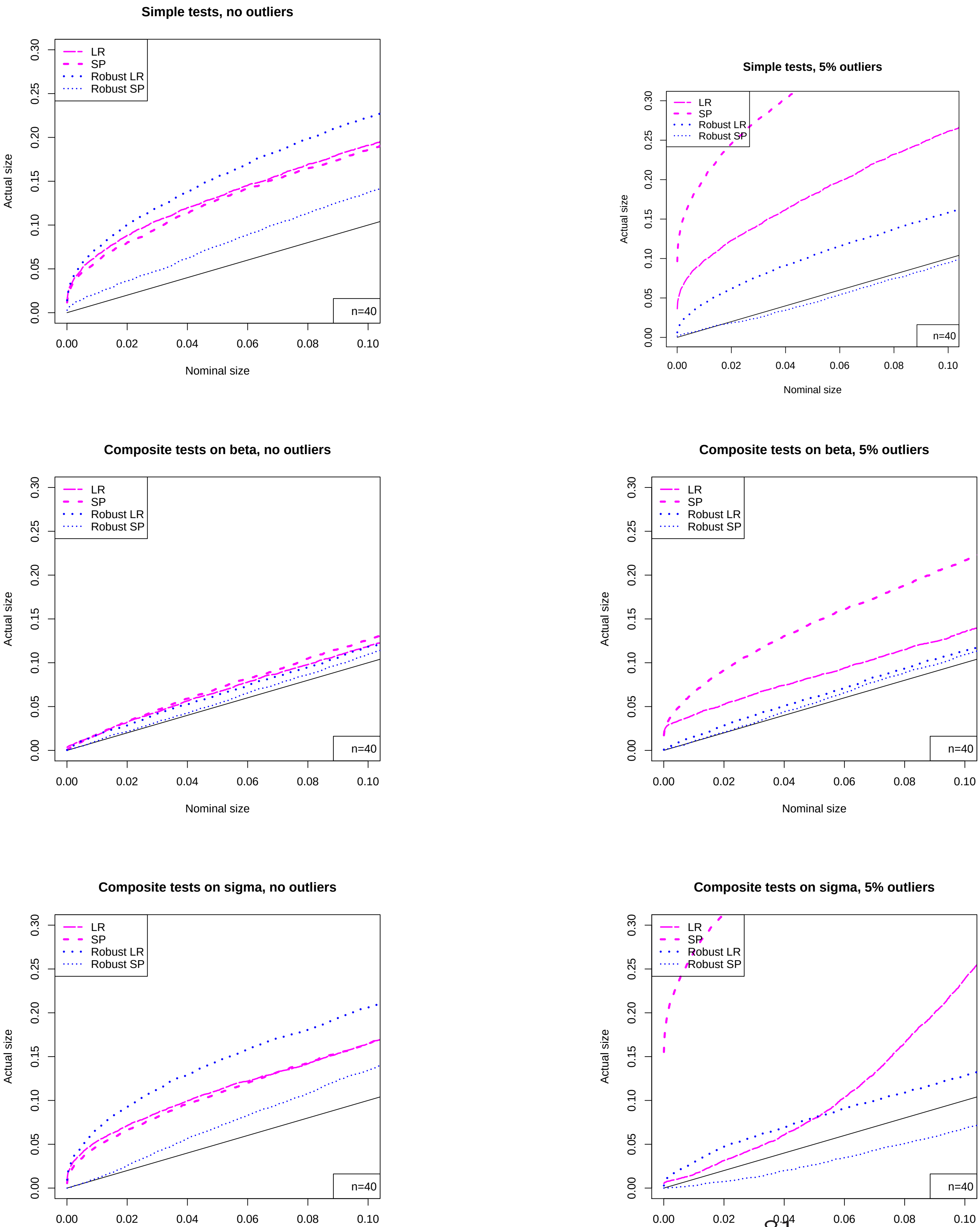
where $\epsilon_t \sim N(0, 1)$.



Schematic illustration of the robust indirect inference algorithm

Genton & Ronchetti (2003) *JASA*

Actual size vs Nominal size
of simple and composite tests
$\beta = -0.05, \;\; \sigma = 0.2, \;\; n = 40.$


Simple tests, no outliers
Simple tests, 5% outliers
Composite tests on beta, no outliers
Composite tests on beta, 5% outliers
Composite tests on sigma, no outliers
Composite tests on sigma, 5% outliers
LR
SP
Robust LR
Robust SP
n=40
Actual size
Nominal size
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.02
0.04
0.06
0.08

♦ Conclusions

- Small Sample Asymptotics can be applied to a large class of models. Its application is particularly useful in robust statistics.

- SSA provides high accuracy for moderate to small sample sizes in the tails of the distribution.

- The saddlepoint test is more accurate than classical tests. It can be applied whenever a score function is available.

- Other applications: quantile regression, composite likelihood, time series in frequency domain.

- Potential use of SSA from an analytic point of view.